# Big data : une mine d'or, mais à quel prix pour notre vie privée ?

Delforge, Antoine

*Published in:*
Revue Nouvelle

*Publication date:*
2017

*Document Version*
le PDF de l'éditeur

# *Big data* : une mine d'or, mais à quel prix pour notre vie privée ?

*Antoine Delforge*

**Le *big data* est un phénomène qui se caractérise par le volume de données, la vitesse d'analyse nécessaire, la gestion de variété de données, la véracité des données qu'il s'agit d'établir et la maitrise des modes de visualisation. Devenu un outil indispensable des campagnes électorales et des sociétés d'assurance, il pose cependant un grand nombre de questions éthiques. Un cadre juridique a donc été défini au niveau européen. Il reste cependant à fournir un lourd travail de sensibilisation.**

Depuis maintenant une petite dizaine d'années, une quantité inimaginable de données sont créées chaque jour, chaque heure, chaque seconde, et ce sans que la plupart des gens le sachent. Ainsi, à chaque fois que vous utilisez votre ordinateur, votre navigateur Internet, votre smartphone, vos objets connectés (montre, voiture...), vous générez inévitablement des données. Quand on sait qu'il y a aujourd'hui 3,9 milliards de personnes connectées à Internet[1] et qu'on estime à plus de 50 milliards le nombre d'objets connectés en 2020 (contre 12 milliards actuellement)[2], vous imaginez vite la quantité d'informations qui circulent constamment.

Avec la généralisation d'un internet à haut débit, 90 % des données numériques ont été produites durant les deux dernières années.

Cette masse de données est appelée « *big data* », ce qui peut être traduit par « données massives » ou encore « mégadonnées ». Cette expression fait référence à des ensembles de données que les outils classiques d'analyse ne sont plus à même de gérer efficacement soit parce que le volume est trop important, soit parce que ces données sont trop brutes (non structurées[3]).

Ces données constituent une mine d'or pour qui les détient et sait les étudier pour leur donner toute leur valeur. Et c'est là que les problèmes surviennent. En effet, pour tirer des informations de ces données brutes, des algorithmes capables de donner du sens à ce flux quasi infini de 0 et de 1 sont nécessaires et le développement de pareil algorithme reste fort onéreux.

Actuellement, la plupart de ces données ne sont pas exploitées, faute de connaissances, de moyens... Certains prétendent que seulement 1 % de celles-ci serait réellement utilisé à sa juste valeur. Les entreprises du numérique analysent pour la plupart les données que chaque internaute produit en utilisant leurs services, mais les entreprises « classiques » aussi ont des stocks de données insoupçonnés qui, une fois exploités, amélioreraient leur productivité. Sur la base de ces monitorings, une étude d'efficacité permet de repenser les processus de fabrication pour optimiser l'ensemble de la production. C'est ce qu'on appelle notamment les « *smart factories* » ou l'industrie 4.0.

Pour bien comprendre ce phénomène du *big data*, il faut se pencher sur les « 5 V » qui le caractérisent : volume, vélocité, variété, véracité, visibilité.

Volume : comme son nom l'indique, le *big data* se caractérise par une énorme quantité de données. À titre d'illustration, 10 millions de DVD Blu-ray pourraient être remplis par les données générées chaque jour.

Vélocité : pour prendre des décisions en conséquence des résultats obtenus grâce aux analyses *big data*, les données doivent être analysées très rapidement, voire en temps réel. Pensez notamment au marché financier[4] où chaque seconde compte.

Variété : si avant l'analyse de données ne pouvait se faire que sur des jeux de données préformatés, desquels il était facile de tirer des statistiques, maintenant des algorithmes avancés peuvent eux-mêmes formater des données de sources très variées : des tweets, des photos, des likes, des vidéos...

Véracité : avec une pareille quantité de données, venant de sources très différentes et difficilement vérifiables, connaitre la véracité de ces données pour en tirer des résultats probants est devenu un enjeu économique majeur.

Visibilité : ce dernier V est apparu plus récemment. Si obtenir des informations de cette masse de données informes reste une chose déjà peu aisée, réussir à les rendre compréhensibles par un chef d'entreprise, par exemple, en est encore une autre. Le *big data* se développant dans des secteurs non spécialisés dans l'étude de données, il est devenu nécessaire de créer des logiciels permettant d'élaborer des graphiques en 2 ou 3D illustrant les résultats obtenus via une analyse *big data* pour ainsi pouvoir être directement utilisé, par la direction d'une entreprise notamment.

Ces quelques explications devraient suffire à vous convaincre du potentiel qu'offre le *big data* : développement économique, apparition de nouveaux services, évolution en matière de recherche (médicale et autres)... et tant d'autres choses insoupçonnées tant on est encore au début d'une révolution industrielle qui viendra toucher tous les secteurs d'activités, sans exception.

## Quelques exemples parlants

« Très bien », me direz-vous, « si c'est bon pour l'économie ! ». « Mais ces données peuvent être des informations personnelles qui en révèlent beaucoup sur vous », vous répondrai-je. Certains prétendront qu'ils n'ont rien à cacher et que dès lors, ils ne sont pas opposés

1 | Plus de 80 % des gens en Europe et Amérique du Nord, chiffres venant du rapport « ICT facts and figures 2016 » réalisé par l'Union internationale des télécommunications, disponible sur http://itu.int.

2 | Ericsson White Paper, « More than 50 Billion Connected Devices », 2011, disponible sur http://bit.ly/2omprww.

3 | C'est-à-dire que les données ne sont pas organisées sur la base de formats prédéfinis (exemple une page web, une image...). À l'inverse, des données structurées sont préformatées afin d'être analysées facilement à l'aide d'un programme informatique, comme un tableur Excel.

4 | Les analyses *big data* sont très utilisées par le secteur financier. Voir LesEchos.fr (S. Rolland), « Comment le *big data* s'impose dans la gestion d'actifs », http://bit.ly/2kcR9dz.

à l'exploitation par une entreprise de leurs données à caractère personnel[5].

Le *big data* est aussi utilisé par de nombreux États, et là paradoxalement les gens sont souvent plus rétifs à partager des informations les concernant. Aujourd'hui, on divulgue plus facilement des informations très personnelles sur Facebook qu'à l'État, alors que l'utilisation qui en est faite par Facebook, pour ne citer que lui, demeure beaucoup plus obscure.

L'analyse de ces données personnelles a des influences quasi quotidiennes sur votre vie sans même que vous le sachiez. En voici quelques exemples très concrets.

### *Le* big data *dans les campagnes électorales : un outil devenu indispensable*

De tout temps, la publicité politique a existé. Quand elle est ciblée, elle devient beaucoup plus efficace que la publicité classique. Le monde politique s'est donc vite intéressé au *big data* comme outil de campagne. Le *big data* permet en effet de très bien connaitre les opinions de chacun. Pour cela, les « data analysts » se fondent sur les données récoltées par des entreprises de marketing, les données publiquement accessibles sur Internet... pour prédire l'opinion des futurs électeurs et même l'influence d'une nouvelle mesure sur la répartition des votes.

Ainsi dès 2002, Mitt Romney, candidat au poste de gouverneur du Massachusetts, a été dans les premiers à utiliser le *big data* afin de cibler davantage les potentiels donateurs pour sa campagne.

Barack Obama a également suivi cette nouvelle tendance, allant même jusqu'à dépenser près de 30 millions de dollars dans l'achat de fichiers auprès notamment de Data Brokers (courtier de données[6]). En 2005, Nicolas Sarkozy a également employé ce genre de technique pour obtenir des listings afin d'envoyer de manière massive des courriels à ses électeurs potentiels qu'il avait réussi à cibler en fonction de leurs « traces numériques ». Ces envois ont été si massifs qu'à l'époque l'expression « Sarkospam » était apparue.

Durant les dernières élections présidentielles américaines tant les Républicains que les Démocrates firent usage d'un service tel que Nation builder, service tout-en-un pour les campagnes électorales. Ce type de service regroupe divers outils d'analyse et de gestion afin de maximiser une campagne électorale sur la base de données personnelles accessibles sur les réseaux sociaux ou acquises d'autres manières (achat, récolte d'informations par les militants...).

L'élection présidentielle française de 2017 n'échappe pas à la règle puisque presque tous les candidats à la primaire de la droite basent leur campagne sur Nation Builder. Mais ce ne sont pas les seuls, Jean-Luc Mélenchon ou encore Emmanuel Macron sont également devenus adeptes de ce type de service.

Tous accros au *big data* ? Pourquoi ? À cause de son efficacité. Ces logiciels permettent d'identifier les sensibilités de chaque personne en fonction des informations la concernant qui circulent sur Internet. Grâce à cela, ces candidats peuvent organiser des campagnes publicitaires en ligne très ciblées (et souvent moins chères que les campagnes non numériques), peuvent capter l'opinion des gens sur certaines thématiques à travers des messages postés sur Facebook ou Twitter. Ainsi, ils adaptent leur discours en fonction du public qu'ils ont en face d'eux. L'application KnockIn, utilisée par les équipes de Nicolas Sarkozy, en est le parfait exemple. En analysant les informations disponibles sur Internet (Facebook et Twitter principalement), cette application catégorise les futurs électeurs en fonction de leur tendance politique et des sujets qui leur sont chers... et tout cela uniquement en regardant leurs activités sur les réseaux sociaux. Ainsi il suffit aux militants qui font du porte à porte de consulter le profil des personnes qu'ils vont démarcher et ainsi formater leur discours pour l'adapter le mieux possible aux aspirations de leurs interlocuteurs.

Avec cet exemple, on constate que la récolte de données personnelles par des organisations politiques peut être vue comme une bonne chose pour prendre le pouls de la population, mais cela incite aussi et surtout les candidats à proposer ce que les gens veulent entendre et pas à soumettre un véritable projet politique qu'ils ont eux-mêmes construit. On passe ainsi d'une politique de l'offre à une politique de la demande.

À une époque où les sondages se trompent de plus en plus souvent, les partisans du *big data* prétendent que l'analyse des traces numériques (like, partage, nombre de vues...) pourrait remplacer les sondages classiques qui seraient par nature biaisés par les questions soumises, les réponses proposées, le manque de sincérité... Il est vrai que seule une analyse *big data* avait vu la victoire de François Fillon à la primaire de la droite[7]. Une autre étude, basée quant à elle sur le nombre de fois où le nom des candidats a été recherché sur Google et cité sur Twitter, aurait également anticipé le résultat de la primaire de la gauche[8]. L'avenir nous dira si ces nouvelles méthodes sont plus crédibles que les sondages d'opinion[9].

### *Le* big data *et les assurances : deux mondes faits pour s'entendre*

Le secteur des assurances a été l'un des premiers à s'intéresser au *big data* et c'est logique. En effet, les assureurs cherchent à connaitre le mieux possible leurs assurés pour pouvoir proposer une prime qui correspond à leurs différents profils et donc aux différents niveaux de risque qu'ils font peser sur la compagnie d'assurance.

À défaut, un phénomène d'anti-sélection va se créer. De fait, les bons clients (à faibles risques) se verront proposer des primes trop élevées et se tourneront vers un concurrent dont le montant de la prime correspondra mieux à leur profil. Il ne restera alors que les mauvais clients dont le trop faible niveau des primes ne correspond plus au niveau des risques. L'équilibre entre bons et mauvais clients est rompu, ce qui pose alors de gros problèmes aux compagnies d'assurance.

On comprend donc très bien pourquoi les assureurs font leur maximum pour tout savoir de leurs clients. Historiquement, les assureurs réussissaient à dresser un profil sur la base de formulaires que le candidat à l'assurance se devait de remplir loyalement.

L'augmentation du nombre d'objets connectés (phénomène que l'on appelle « Internet des objets » ou IoT, « *internet of things* ») devient une nouvelle chance pour eux de cerner encore plus précisément leurs assurés. En effet, si ces compagnies réussissaient à avoir accès aux informations de santé générées par les montres ou bracelets connectés, elles

5 | Les exemples d'utilisation de ces données qui vont suivre pourraient faire changer d'avis ces personnes.

6 | Cette activité de *data brokers* consiste à acheter diverses bases de données sur des personnes. On croise alors ces différentes bases afin d'obtenir une information plus complète sur chaque personne concernée. Enfin, le tout est revendu à différentes sociétés qui souhaitent avoir accès à ces profils.

7 | V. Queffelec (CEO d'Euromediations), « Seul le *big data* a annoncé Fillon gagnant », disponible sur http://bit.ly/2pQLHiA.

8 | LePoint.fr (S. Edelson), « Primaire de la gauche : quand Twitter et Google permettent de prédire les résultats », disponible sur http://bit.ly/2pR3p5v.

9 | Pour vérifier cela par vous-même, la société ayant vu la victoire de Fillon pronostiquait le 15 janvier Fr. Fillon et E. Macron en tête au premier tour de la présidentielle française, tout en rappelant que plusieurs éléments majeurs pouvaient encore modifier ces résultats. Voir V. Queffelec (CEO d'Euromediations), « Fillon, Macron, Le Pen ? », disponible sur http://bit.ly/2oNaQeS.

pourraient proposer une assurance vie dont la prime n'est plus calculée à partir d'un profil établi sur la base de statistiques, mais propre à chaque personne. Cette prime serait alors fixée en fonction du rythme cardiaque de l'assuré, de son activité physique, de son stress, des lieux qu'il fréquente (tout cela étant des données que peut récolter un bracelet connecté) et de toute autre information qui aurait une influence sur son espérance de vie (facteur qui détermine le montant des primes d'assurance vie).

Un autre exemple, les assurances RC auto. Dans les voitures connectées, il y a des dizaines de capteurs qui enregistrent différents paramètres (nombre de kilomètres parcourus, vitesse moyenne, manière de freiner...) permettant finalement de connaitre le type de comportement du conducteur. Cette information, si elle était connue des assureurs, permettrait de fixer une prime au cas par cas en fonction de la manière de conduire du demandeur d'assurance. Le conducteur qui a un style de conduite très sportif, qui roule vite et freine tard verra sa prime augmenter, là où une personne au style plus coulé verra la sienne diminuer, car son style est jugé moins accidentogène.

Où est le danger dans ce genre de pratique ? La personne en bonne santé et qui conduit bien serait sans doute d'accord que son assureur ait accès à ses données si cela peut faire diminuer ses primes d'assurance. Certes, les primes peuvent diminuer, mais cela veut dire que si la personne arrête, par exemple, subitement de faire du sport, l'assureur va remarquer un changement dans son style de vie grâce au bracelet connecté qu'elle s'est engagée à porter, et ce changement pourrait avoir comme conséquence de faire réaugmenter les primes d'une assurance vie. De même, si elle décide du jour au lendemain de se mettre à conduire comme un pilote de Formule 1, sa voiture transmettra ce brusque changement à l'assureur qui le répercutera alors sur le montant de sa prime.

D'autres problèmes peuvent apparaitre dans l'hypothèse où ce phénomène venait à se développer.

L'ultra-personnalisation tend à faire disparaitre l'aléa qui demeure l'un des principes clés en matière d'assurance. Quand le *big data* permet de pronostiquer avec une quasi-certitude l'avenir d'un assuré, où est encore le risque pour l'assureur ?

Un autre problème pourrait venir du fait que les assurances vont plus que probablement inciter les gens à partager leurs informations personnelles, quitte à ce que la montre connectée soit offerte à la conclusion de chaque assurance complémentaire santé, comme l'a testé pendant un moment AXA France. Si ce genre d'offre réussissait à convaincre une grande partie des assurés, la partie réfractaire à ces traqueurs d'activités relayant des données aux assureurs verrait probablement à terme sa prime augmenter. En effet, les assureurs auront face à eux un groupe plus risqué, puisque moins transparent que les autres. De plus, les personnes n'ayant pas voulu contracter une assurance liée à un objet connecté deviendront à priori des gens qui ont quelque chose à cacher à leur assureur.

De plus, cette ultra-segmentation aura pour conséquence que les primes vont être de plus en plus variables[10] et que les personnes avec un profil considéré à haut risque se trouveront devant un danger de non-assurabilité à la suite de l'augmentation importante de leur prime qu'il leur sera devenu impossible de payer.

10 | Le montant variera, mais la période entre deux réévaluations du niveau de risque pourrait aussi fortement diminuer. Techniquement, on peut imaginer aller jusqu'à une réévaluation quotidienne, voire même en temps réel.





En cas de développement de pareilles offres, seule une intervention législative peut prévenir ces différents problèmes.

### *Avec le* big data, *chacun sa vie, chacun son web*

Internet est un monde ouvert où toute personne connectée peut aller consulter presque tout ce qu'elle veut sans aucun contrôle. En tout cas, c'est ce que l'utilisateur non averti pense.

En réalité, il n'en est rien. Malgré cette illusion de liberté, chaque site internet peut décider ce qu'il veut que chaque visiteur voit, deux visiteurs d'un même site n'ont donc probablement pas exactement le même contenu devant les yeux ; et tout cela grâce aux « cookies ». En informatique, les cookies sont de petits fichiers semblables à des fichiers textes déposés par un site internet sur l'ordinateur (ou le smartphone, la tablette...) de la personne qui consulte cette page. Ce fichier contient plusieurs informations, dont notamment un numéro permettant de reconnaitre cet appareil la prochaine fois qu'il retourne sur le site. Voilà comment expliquer que, par exemple, on ne choisit généralement que la première fois la langue d'une page web. La fois suivante, grâce au cookie, le site internet sait qu'il doit envoyer la page en français pour cette personne-là.

Ces cookies permettent donc de suivre une personne sur Internet, les différents sites s'échangeant massivement les cookies qu'ils ont déjà disséminés afin de savoir ce qu'a consulté telle personne sur tel autre site.

Ce mécanisme simple permet ainsi à Amazon, pour ne parler que de lui, de faire réapparaitre sur son site, contre rémunération, des annonces consultées plusieurs jours auparavant sur des sites de presse, sur Facebook...

Amazon est loin d'être le seul à utiliser ce genre de techniques pour proposer du contenu personnalisé. Tous les sites proposant des systèmes de recommandation fonctionnent de la même manière, sur la base de l'historique de navigation de chaque utilisateur recréé à partir de ces cookies. Certains sites couplent ces données avec d'autres éléments qui peuvent influencer les choix et les envies des consommateurs potentiels.

Dans ce domaine, Netflix a un des algorithmes de recommandation les plus évolués ; de sorte qu'en fonction de la météo, de l'heure, de votre historique, du type de contenu que vous regardez (thème, durée, avis), il vous est suggéré un contenu qui correspond à votre envie du moment. Netflix va plus loin encore en se basant sur ces informations pour produire des séries ou des films qui correspondent à ce que souhaite voir son public, ce qui n'est pas sans poser question sur le risque de manque de diversité que cela pourrait créer.

Sur Facebook, les publications suggérées sont choisies par un algorithme qui définit ce qui est susceptible de vous intéresser en fonction de vos likes, des pages consultées à partir de Facebook..., en fonction de vos gouts. Il tient également compte du comportement de vos amis sur le réseau social. Ce qui à première vue peut sembler pratique provoque une certaine isolation où l'on ne voit en réalité qu'une partie de ce qui circule sur Internet. C'est l'effet « bulle de filtres », car chaque internaute reste limité malgré lui à un univers pensé et créé sur mesure. Cette bulle a en plus tendance à s'autoalimenter à tel point que chaque personne dans cette bulle ne perçoit le monde que par le prisme propre à cette bulle[11].

Google, grâce à l'historique des recherches, fait de même, de sorte que les résultats d'une recherche peuvent varier

11 | Pour s'en convaincre, voir notamment un montage réalisé par le *Wall Street Journal* qui met en perspective deux fils d'actualités Facebook : un orienté républicain et un orienté démocrate, http://bit.ly/2axcSbb.

très fortement d'un individu à l'autre, d'un ordinateur à l'autre pour être plus précis. Ainsi, en tapant Égypte dans la célèbre barre de Google certains verront apparaitre des sites de voyage, là où d'autres seront redirigés vers des articles parlant des manifestations dans le pays[12].

Percer cette bulle demeure compliqué, mais effacer régulièrement ses cookies reste une des solutions les plus faciles à mettre en place[13].

### Big data : *la chasse à la fraude*

Le secteur public a également vu dans le *big data* une belle opportunité, notamment dans la chasse à la fraude sociale et fiscale.

L'administration fiscale belge utilise depuis plusieurs années déjà des méthodes d'analyse *big data* (plan dataminig lancé par le SPF Finances) croisant de multiples informations disponibles au sein des différentes administrations afin de créer des profils de fraudeur. Sur la base de ces profils, les contrôles sont orientés vers des personnes considérées comme plus probablement fraudeuses.

Le croisement de données se fait aussi avec des informations venant des fournisseurs et gestionnaires d'eau, de gaz et d'électricité. S'il existe une différence significative entre la consommation réelle et la consommation moyenne d'un même type de ménage que celui habitant officiellement à cette adresse, l'inspection sociale est alertée et sera alors chargée de vérifier qu'il n'y a pas de fraude sociale, une domiciliation fictive par exemple, afin de toucher des allocations en tant que personne isolée alors que, dans les faits, le bénéficiaire est en cohabitation.

Le *big data* permet également de détecter plus facilement les montages fiscaux (carrousel TVA...) où la détection de schéma organisationnel se fait beaucoup plus rapidement et efficacement.

Ces applications *big data* ont cependant tendance à cibler toujours les mêmes des profils et cela a déjà provoqué des contrôles fiscaux à répétition.

### Protection des données personnelles : un cadre juridique mis à jour

À travers ces différents exemples, on comprend rapidement que l'exploitation à grande échelle de données personnelles peut avoir de graves conséquences si cette exploitation se pratique sans aucune règlementation. Imaginez que n'importe quelle entreprise puisse récolter ces données, à votre insu et surtout sans votre consentement, et les vendre à qui les veut ! Plus aucune vie privée sur Internet. C'est pourquoi dès 1995, l'Union européenne a encadré les traitements de données personnelles.

Le monde a beaucoup évolué depuis ce moment-là. L'informatique, et plus spécifiquement Internet, n'a cessé de prendre de plus en plus de place dans notre vie. On est passé de pages web purement passives à des pages web interactives où des données sont échangées en permanence. Google, Amazon, Facebook, Apple, Microsoft (regroupé sous l'acronyme Gafam) sont devenus en vingt ans non pas uniquement des sociétés d'informatique dont les services sont maintenant inévitables, mais des sociétés qui collectent un maximum de données sur lesquelles reposent une importante partie de leur modèle économique.

Tous ces changements ont donc poussé l'Union européenne à adapter sa législation à ce qu'on appelle le web 2.0.

Ce processus, commencé en 2012, prit près de quatre ans pour aboutir à l'adoption du nouveau Règlement général sur la protection des données qui n'entrera

12 | Exemple tiré d'une conférence sur le sujet donnée par Eli Pariser, http://bit.ly/2ommVqa.

13 | Les bloquer n'est pas possible parce qu'ils sont devenus indispensables à de nombreux sites. Les bloquer revient alors à rendre ces pages inaccessibles.





en vigueur qu'en mai 2018 pour laisser le temps à tout le monde de s'adapter.

### *Quelles données sont protégées ?*

La législation relative à la protection des données à caractère personnel vise à règlementer la récolte et l'usage de ces données.

On parle de données à caractère personnel face à une information qui peut être rattachée à une personne identifiée ou identifiable. Ainsi, un numéro de compte, un cookie... sont des moyens qui permettent de discriminer une personne par rapport à une autre et dès lors toutes les informations liées à cet identifiant sont considérées comme des données à caractère personnel. Ces informations peuvent être de toutes natures (texte, vidéo, image...). Cela peut aller de votre nom, votre adresse ou votre âge, à vos historiques web en passant par vos habitudes alimentaires, sportives... tout ce qui se rapporte à une personne pouvant être identifiée sans trop de difficultés.

Ce nouveau Règlement s'applique à tous les services destinés à un public européen, de sorte que toutes les entreprises étrangères (le Gafam et autres) sont tenues de respecter cette législation quand elles traitent des données relatives à des Européens. À défaut, elles risquent des amendes très élevées pouvant grimper jusqu'à 4 % de leur chiffre d'affaires annuel mondial.

### *Comment sont-elles protégées ?*

La règlementation en la matière se compose de cinq grands principes, desquels découle une série de droits pour les personnes dont les données sont traitées (appelées « personnes concernées » par la suite).

### *Cinq grands principes*

1. Principe de finalité : les données doivent être récoltées dans un but spécifique et ne peuvent être utilisées qu'à cette fin. Cette finalité est annoncée dès l'obtention de ces informations. Ainsi, la finalité d'une collecte d'adresses courriels pourrait être d'envoyer ultérieurement des publicités.
2. Principe de licéité : des données personnelles ne peuvent pas être récoltées et/ou utilisées sans le consentement de la personne concernée[14]. Ce consentement ne portera que sur la finalité annoncée. Dans certains cas, un consentement n'est pas nécessaire. Tel est le cas quand ces données sont traitées par l'administration, par exemple, dans le cadre d'une mission d'intérêt public ou quand ces données doivent être récoltées pour pouvoir fournir un service demandé par la personne.
3. Principe de minimalisation : seules les données nécessaires à la finalité déclarée peuvent être traitées. Ce principe s'applique tant à la quantité de données qu'à leur durée de conservation. Dès lors, régulièrement, certaines données devenues inutiles doivent être effacées, de sorte que logiquement les fournisseurs de service n'ont pas à conserver les données de clients dont le contrat s'est terminé il y a plusieurs années. En principe, ces informations doivent être supprimées de leur listing.
4. Principe de transparence : l'entreprise ou le service public qui traite des données personnelles doit être transparent sur ce qu'il fait des données et comment il le fait. Il existe une obligation d'information. Cette information comporte notamment la liste des entités à qui les données vont être communiquées, la durée de conservation, l'explication des droits qu'a la personne concernée[15].

14 | Exemple : les bandeaux cookies qui apparaissent sur de nombreuses pages web pour demander l'autorisation de traiter vos données via un cookie.

15 | Les principaux sont expliqués juste après.

5. Principe de sécurité : le responsable du traitement de données personnelles est obligé d'assurer un niveau de sécurité adapté aux risques que pourraient occasionner une fuite, une suppression ou une altération des données qu'il gère. Tous ces principes ne serviraient, en effet, à rien si le premier pirate informatique venu se servait comme il l'entend. Ce niveau de sécurité variera donc en fonction du volume de données et de leur nature. Une société comme Facebook doit dès lors avoir un niveau de sécurité très élevé. Elle détient de fait énormément de données, parfois très sensibles, pouvant être dommageables si elles tombent entre de mauvaises mains.

### *Vos droits en cas de traitement de vos données*

En cas de traitement de données personnelles, les personnes concernées ont différents droits qu'elles peuvent faire valoir face au responsable de ce traitement[16]. Mais, face à certains pouvoirs publics, ces droits ne peuvent être invoqués.

Vous avez d'abord le droit de savoir si des données vous concernant sont en possession de l'organisation que vous interrogez et, dans l'affirmative, lesquelles. Le nouveau Règlement autorisera également d'en exiger une copie.

Ensuite, vous pouvez réclamer qu'une information erronée soit corrigée ou mise à jour, voire supprimée. Cela permet d'éviter que cette erreur entraine un profilage inapproprié par exemple, ce qui peut avoir de lourdes conséquences puisque de ce profilage va dépendre une série de choses[17].

De plus, vous êtes en droit de demander, dans certains cas, à ce que vos données personnelles soient effacées (droit à l'oubli). Ainsi pour prendre un cas fréquent, si une page peu flatteuse vous concernant est référencée sur Google et que vous souhaitez éviter que l'on y accède en tapant votre nom dans ce célèbre moteur de recherche, vous pouvez exiger de Google qu'il vous déréférence[18], si cette demande est légitime.

Enfin, le nouveau Règlement a introduit le droit à la portabilité. En 2018, vous pourrez réclamer que vos données personnelles vous soient envoyées de manière à ce que vous puissiez les réinsérer dans un autre service du même genre. Pour éviter de rester emprisonné dans un service particulier, vous pourrez plus facilement changer de réseau social, de service de Cloud ou messagerie web, sans voir toutes vos données perdues.

### *Flux de données hors Union européenne*

Si l'usage qui peut être fait de données personnelles est bien règlementé en Europe, que se passe-t-il quand ces données sont transférées hors de l'Union européenne (ce qui est très souvent le cas) ? Comment contrôler ce qui est fait des données rapatriées par Facebook sur ses serveurs situés dans la Silicon Valley ?

La réponse est simple. En principe, les données personnelles de citoyens européens ne peuvent quitter l'Europe. Seuls les transferts vers des pays où un niveau de protection des données adéquat existe sont autorisés[19]. Tel est, par exemple, le cas pour le Canada pour le secteur privé. Dès lors, des transferts de données personnelles venant d'Europe peuvent être effectués vers des serveurs d'entreprises privées canadiennes. La règlementation sur la protection des données ne freinera pas les rapprochements économiques entre l'Union européenne et le Canada.

Cependant quand on parle de pays nord-américains et de données, tout le monde pense à la NSA et à l'affaire Snowden. Pour être bref, avant cette affaire, l'Union européenne avait négocié avec les États-Unis (pour Facebook...) le « Safe Harbor », un cadre juridique dans lequel les États-Unis s'engageaient à respecter les principes de la législation européenne en matière de protection des données. Avec la révélation du programme « Prism » de la NSA par Edward Snowden, la cour de Justice de l'Union européenne a considéré que le « Safe Harbor » n'apportait plus les garanties suffisantes pour les citoyens européens. Cette décision poussa l'UE à renégocier avec les États-Unis un nouvel accord censé être plus protecteur, nommé « Privacy shield ». Cet accord n'est toujours pas suffisant pour de nombreux observateurs.

## En conclusion

En écrivant ce papier, nous comprenons bien que le mariage *big data* et données personnelles peut en effrayer plus d'un, qui pourrait devenir paranoïaque à chaque fois qu'il surfe sur Internet. Tel n'est pas le but de notre propos.

Le *big data* combiné avec des données personnelles reste avant tout une opportunité de progrès pour le moment sous-estimée. Cela peut faire avancer rapidement la recherche médicale, améliorer grandement l'efficacité des entreprises et des services publics...

Il est vrai que les sociétés telles que Google ou Facebook génèrent énormément d'argent sur la base de nos données personnelles, il est également vrai que nos vies dépendent parfois de la manière dont elles vont utiliser ces données.

Mais, par ailleurs, ces services sont proposés sans contrepartie, si ce n'est le prix que valent nos données évidemment, que nous cédons plus ou moins consciemment. Le rapport de force entre ces sociétés et leurs utilisateurs reste fortement à l'avantage des premières.

Et là entre en jeu la règlementation sur la protection des données personnelles. Elle permet de rééquilibrer les choses, du moins en partie. Cela rend les processus de traitement de données personnelles plus transparents ce qui améliore la prise de conscience collective. Si tout le monde était sensibilisé à ces questions, chacun serait à même de choisir en connaissance de cause ce qu'il veut divulguer (autodétermination informationnelle) et quelles conséquences cela peut avoir pour lui.

16 | Ne sont repris ici que les droits les plus importants.
17 | Voir les exemples cités précédemment.
18 | Pour entamer cette procédure : http://bit.ly/TZh0rH.
19 | Il est également possible pour les entreprises intéressées de recréer contractuellement le même cadre juridique qu'en Europe et ainsi d'assurer que des données quittant l'Europe soient tout aussi bien protégées que si elles restaient sur le vieux continent.